Problem 1

Let

$$co\mathcal{X} = \{x : x = \sum_i \alpha_i x_i, \alpha_i \geq 0, \sum_i \alpha_i = 1\}$$

$$co\mathcal{Y} = \{y : y = \sum_j \alpha_j y_j, \alpha_j \geq 0, \sum_j \alpha_j = 1\}$$

be two convex hulls for $\mathcal{X} = \{x_i\}_{i=1}^N$ resp. $\mathcal{Y} = \{y_j\}_{j=1}^M$.
Prove : If the two convex hulls intersect $\Rightarrow \mathcal{X}$ and $\mathcal{Y}$ are not linearly separable.
Solution.
    Proof by contradiction. Assume simultaneously that the two convex hulls intersect and that the two sets are linearly separable.
$co\mathcal{X} \cap co\mathcal{Y} = \{z : z \in co\mathcal{X} \wedge z \in co\mathcal{Y}\}$, i.e.

$$\begin{cases} z = \sum_i \alpha_i x_i \\ z = \sum_j \alpha_j y_j \end{cases}$$

By second assumption, we have

$$\begin{cases} w^\mathsf{T} z + w_0 > 0 \\ w^\mathsf{T} z + w_0 < 0 \end{cases}$$

Putting it together we get

$$\begin{cases} z = \sum_i \alpha_i x_i \\ z = \sum_j \alpha_j y_j \end{cases}$$
$$\begin{cases} w^\mathsf{T} z = \sum_i \alpha_i w^\mathsf{T} x_i \\ w^\mathsf{T} z = \sum_j \alpha_j w^\mathsf{T} y_j \end{cases}$$
$$\begin{cases} w^\mathsf{T} z = \sum_i \alpha_i w^\mathsf{T} x_i > -w_0 \sum_i \alpha_i = -w_0 \\ w^\mathsf{T} z = \sum_j \alpha_j w^\mathsf{T} y_j < -w_0 \sum_j \alpha_j = -w_0 \end{cases} \perp$$

(Second assumption can not be true, thus the two sets are not linearly separable.)

Problem 2

First, let the decision boundary $w^\mathsf{T} x = 0$ separate the datapoints correctly, i.e. datapoints for which $w^\mathsf{T} x > 0$ lie on one side of the hyperplane and those for which $w^\mathsf{T} x < 0$ lie on the other side. Thus the logistic regression model is correctly classifying each datapoint with probability p of at least $\frac{1}{2}$ (because $\sigma(\cdot) = \frac{1}{2}$ at zero). Now if $||w|| \to \infty$, the datapoints are still classified correctly but the probability $p \to 1$ and $\sigma(\cdot)$ becomes the step function. Assuming labels $\{1, -1\}$, the binary cross-entropy becomes $\sum_{i=1}^N \ln(1 + \exp(-y_i \mathbf{w}^T \mathbf{x}_i))$. We can clearly see that for correctly classified datapoints the loss is zero.

$||w|| \to \infty$ is undesirable because of overfitting but we can prevent it by adding weights regularization to the objective $E(w)$ to penalize large weights, e.g. $\lambda ||w||_2^2$.

Problem 3

We observe that for class of blue crosses and class of black circles, one of the features is positive and the other negative respectively both features are either positive or negative. Thus a possible basis function $\phi(x_1, x_2) = x_1 \cdot x_2$. Now the two classes become linearly separable, since all blue crosses lie on the real axis left of zero and all black circles lie right of zero.

Problem 4

The decision boundary is by construction defined as $\mathbf{w}^\mathsf{T}\mathbf{x} + w_0 = 0$, where $w = \begin{bmatrix} w_1 & w_2 \end{bmatrix}$, $x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^\mathsf{T}$, thus we need three parameters (including the bias).

$$w_1 x_1 + w_2 x_2 + w_0 = 0$$

$$\begin{cases} 2w_1 = -w_0 \\ 5w_2 = -w_0 \end{cases}$$

$$w_1 = \frac{-w_0}{2}$$

$$w_2 = \frac{-w_0}{5}$$

Possible coefficients must satisfy $\mathbf{w}^\mathsf{T}\mathbf{x} + w_0 = 0$. E.g. for $w_0 = -1$, we get the coefficients $w_1 = \frac{1}{2}, w_2 = \frac{1}{5}$.